Outline (IGNITE):

1. Intro –
   1. What is the problem? – To predict counts on an hourly basis for the last 10 days of every month
   2. What is the goal? – To be able to re-balance the bike counts at each of the stations using our predicted results.
2. Pre-processing –
   1. Plots – Used to analyze the data and determine the important analysis components.
   2. What are the variables used to determine counts?
   3. How do we finally create the response variable?
   4. How do we separate to training and test sets?
3. Models –
   1. What is the procedure we need to follow to predict counts and why?
   2. Regression models applied
      1. Ridge regression
         1. Accuracy and results
      2. Linear Regression
         1. Accuracy and results
      3. Multiple regression
      4. Polynomial Regression
   3. What we read from these models? (Behavior of residuals etc)
4. Re-balancing –
   1. Build a distance matrix of all bike stations in Pittsburgh. According to the distance matrix, final count of bike trip in a station on a day of a user type, and the density plot based on Hour attribute to deal with rebalancing problem.
   2. To retransfer the bike into one station, first, we pike stations from the nearest to farther. For example, if one station (name A) has a negative count, which means bikes out of the station are more than bikes into the station, we select the station from the nearest one, if that one (name B) has a positive value, retransfer the necessary number of bikes from B to A, if it is enough, go on to another station with negative value; if it is not enough, go on to another near station with positive value to retransfer necessary number of bike in to A.
5. Conclusion –
   1. What we achieved?
   2. Which is the best Method?
   3. What were the challenges?
   4. How concrete are the results?